Statistical Analysis Plan
Community Apprentice

Evaluator (institution): Behavioural Insights Team

Principal investigator(s): Matt Barnard

Template last updated: July 2018

PROJECT TITLE	Community Apprentice
DEVELOPER (INSTITUTION)	Envision
EVALUATOR (INSTITUTION)	Behavioural Insights Team
PRINCIPAL INVESTIGATOR(S)	Matt Barnard
TRIAL (CHIEF) STATISTICIAN	Matt Barnard
SAP AUTHOR(S)	Patrick Taylor
TRIAL REGISTRATION NUMBER	ISRCTN27341369
EVALUATION PROTOCOL URL OR HYPERLINK	https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/ … [*Please replace with project protocol URL or hyperlink]*

SAP version history

VERSION	DATE	REASON FOR REVISION
1.0 [*original*]	28 September 2018	

Table of contents

Introduction

This intervention aims to develop participants' sense of self-efficacy, social confidence[1] and teamwork skills by supporting year 10 pupils to design and deliver a fundraising campaign. It combines project work in small teams with mentoring by local business employees.

This will be a two-arm randomised controlled trial, with randomisation at the pupil level. The primary outcome is combined GCSE performance, as indicated by a pupil's Attainment 8 score. The effect of the intervention on five secondary outcomes will also be estimated: Maths GCSE attainment, English GCSE attainment, Self-efficacy, Teamwork, and Social Confidence. Further analysis will also be carried out to estimate the Complier Average Causal Effect, and the effect of the intervention on longitudinal outcomes (progression into further social action,post-16 attainment, progression into higher education, and employment status).[2]

All effects will be estimated for the whole sample and for the subgroup of pupils who were eligible for free school meals (FSM) in the year of the trial and/or the previous six years ('Ever 6 FSM').

Design overview

Please ensure all details are in line with the latest version of the protocol.

Trial type and number of arms		Two-arm, individual-level randomisation
Unit of randomisation		Pupil
Stratification variables (if applicable)		School, FSM status
Primary outcome	variable	Combined GCSE performance
	measure (instrument, scale)	GCSE Attainment 8 score
Secondary outcome(s)	variable(s)	Maths GCSE attainment, English GCSE attainment, Self-efficacy, Teamwork, and Social Confidence
	measure(s) (instrument, scale)	Maths GCSE point score, average of English GCSE point scores (i.e. both English Literature and English Language), New General Self-Efficacy Scale, Teamwork Scale for Youth, Self-Perceived Communication Competence Scale

Follow-up

Changes in the analytical sample have not affected the choices of statistical analyses.

[1] 'Social confidence' is defined by Envision as a young person's confidence in communicating with new people (especially unfamiliar adults) in one-to-one, group and presentation settings.
[2] See 'Additional Analysis' below for further details on this.

Sample size calculations overview

Please ensure all details are in line with the latest version of the protocol.

		Protocol		Randomisation	
		OVERALL	FSM	OVERALL	FSM
MDES		0.185	0.342	0.185	0.295
Pre-test/ post-test correlations	level 1 (pupil)	0.71	0.71	0.71	0.71
	level 2 (class)	NA	NA	NA	NA
	level 3 (school)	NA	NA	NA	NA
Intracluster correlations (ICCs)	level 2 (class)	NA	NA	NA	NA
	level 3 (school)	NA	NA	NA	NA
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		NA	NA	NA	NA
Number of schools	**total**	30	30	30	30
Number of pupils[3]	intervention	390	117	389	154
	control	390	117	371	156
	total	780	234	760	310

Justification for the assumptions in the table above are as follows:

- **Pre-test/post-test correlations:** We assumed that 50% of the variance in primary outcome will be explained by Maths and English KS2 scores. So, the combined R^2 for these baseline covariates is 0.5, giving a correlation coefficient of 0.71. This assumption is based on analysis of the Department for Education Longitudinal Study of Young People in England.[4]
- **Alpha and Power:** These are standard assumptions.
- **One or two-sided test:** A two-sided test is performed to err on the side of caution. There is little existing evidence relating to the effect of interventions of this type on academic attainment, so we cannot assume the direction of any effect that we might observe.
- **Number of schools:** The delivery organisation provided the estimate of 30 schools based on their previous experience of recruiting schools to similar interventions and their staff capacity for this intervention.

[3] All figures are for numbers assigned. The MDESs given in this table assumed that 23% of pupils would drop out of the programme (and so receive no benefit from it) but remain in the sample for analysis.

[4] Anders, J. (2016). Not published. This analysis used a linear model of the relationship between KS2 average (across English, Maths and Science) points score and the capped 8 points score for GCSE and equivalent qualifications. The dependent and independent variables are slightly different in this current trial (the KS2 Science assessment has since been dropped and Attainment 8 did not exist previously), but this is the best available analysis that we have to make this estimation.

- **Number of pupils:** The delivery organisation provided the estimate of 780 pupils, based on their maximum capacity to support 13 pupils per intervention group. Based on previous experience of running similar interventions, they were confident that schools would commit to this number. In the end, 760 pupils were randomised.
- **Free School Meals:** Participating schools were asked to aim to select pupils such that a minimum of 30% were eligible for free school meals. We assumed that this aim would be met based on the delivery organisation's prior experience, and because schools were located in areas where intakes have above average proportions of pupils on free school meals. In the end, 310 pupils eligible for FSM (41% of the total sample) were randomised.

Analysis

The analysis plan is described in the sections that follow. All analyses will be carried out using the statistical software Stata (version 14). See Appendix 1 for the associated Stata code that will be used.

Primary outcome analysis

Outcome

The primary outcome is overall GCSE performance, measured by capped average point scores across eight best GCSEs (Attainment 8).

Analysis

Primary analysis will be intention-to-treat (ITT), in which we test the hypothesis that being assigned a place on the programme has an effect on attainment. Analysis will be carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 \cdot T_i + X_i \cdot \Gamma + \epsilon_i$$

where:

- Y_i is the outcome for overall GCSE performance, measured by capped average point scores across eight best GCSEs (Attainment 8);
- T_i is a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i is a vector of individual-level stratification variables (School ID and EVER6_FSM_P) and the baseline attainment measured through separate KS2 Reading (KS2_READMRK) and Maths (KS2_MATTOTMRK) marks[5]; and
- ϵ_i is the individual-level error term. Standard errors will be corrected for heteroskedasticity.

While point scores are bounded, we assume that the response to the treatment will be locally linear so an OLS will be appropriate.

[5] This is a deviation from the protocol, which suggested the use of English and Maths raw scores. Raw scores are not available from the National Pupil Database, so we are using KS2_READMRK (marks achieved in the English reading test) and KS2_MATTOTMRK (the sum of Maths Paper A, Maths Paper B and mental arithmetic tests). The KS2 English writing test is teacher assessed so will not be included as an indicator of English attainment.

Secondary outcome analysis

Outcomes

Effects will be estimated for five secondary outcomes:

- **English attainment**, as measured by both English Literature and English Language GCSE with point scores combined to create a mean average English score.
- **Maths attainment**, as measured by Maths GCSE point score.
- **Self-efficacy**, as assessed by the "New General Self-Efficacy scale" (NGSE).[6] This is a 7-item self-report survey that contains items relating to pupils' sense of general self-efficacy. Participants are asked to respond on a scale of 1 (strongly disagree) to 5 (strongly agree). In line with the scale developers' own approach[7], the mean average score on the scale for each pupil will be used in the analysis.
- **Teamwork,** as assessed by the "Teamwork Scale for Youth" (TSY).[8] This is an 8-item self-report survey, that contains items relating to working with others. Participants are asked to respond on a scale of 1 (strongly disagree) to 5 (strongly agree). In line with the scale developers' own approach[9], the mean average score on the scale for each pupil will be used in the analysis.
- **Social confidence**, as assessed by the "Self-Perceived Communication Competence Scale" (SPCCS).[10] This is a 12 item self-report survey that measures self-perceived communication in a range of contexts. Participants are asked to rate their competence in 12 situations between 1 and 100. In line with the scale developers' own approach[11], the mean average score on the scale for each pupil will be used in the analysis. This scale is also designed to produce 7 sub-scores that reflect different types of communication competence, which are theoretically relevant to the intervention. If an effect is observed using the average score, as exploratory analysis we will also estimate effects using these sub-scores, which are calculated as follows.
    - Public: The sum of the scores from items 1, 8 and 12; divided by 3.
    - Meeting: The sum of the scores from items 3, 6 and 10; divided by 3.
    - Group: The sum of the scores from items 4, 9 and 11; divided by 3.
    - Dyad: The sum of the scores from items 2, 5 and 7; divided by 3.
    - Stranger: The sum of the scores from items 1, 4, 7 and 10; divided by 4.
    - Acquaintance: The sum of the scores from items 2, 6, 9 and 12; divided by 4.
    - Friend: The sum of the scores from items 3, 5, 8 and 11; divided by 4.
    - Total: The sum of all scores; divided by 12 (i.e. the mean score for the scale).

As suggested in the evaluation protocol, we also considered using the speaking and listening ('Spoken Language') component of GCSE English as an additional measure of social

[6] Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational research methods*, 4(1), 62-83.
[7] Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational research methods*, 4(1), 68.
[8] Lower, L. M., Newman, T. J., & Anderson-Butcher, D. (2015). Validity and reliability of the teamwork scale for youth. *Research on Social Work Practice*.
[9] Lower, L. M., Newman, T. J., & Anderson-Butcher, D. (2015). Validity and reliability of the teamwork scale for youth. *Research on Social Work Practice*, 721.
[10] McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication Research Reports*, 5:2, 108-113.
[11] McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication Research Reports*, 5:2, 111.

confidence. After investigating this measure further, we have decided not to use it because the moderation process is significantly less robust than the exam assessments, the outcome variable is teacher assessed and coarse (Pass, Merit, Distinction, Not Classified)[12], and the variable is not available on the NPD.

Outcomes 3 to 5 in the list above have been collected in a single paper survey (see appendix 2), administered in schools by BIT Research Assistants (RAs). The data was collected after the end of the intervention between 25th April 2018 and end of the school year (i.e. mid July 2018). See Appendix 2 for a copy of the survey. This survey was first piloted with four Y10 pupils prior to administration to ensure the length and wording of the survey was appropriate for this study. As a result of this pilot, the following changes were made to the original survey wording.

Original wording	New wording
NGSE2: When facing difficult tasks, I am certain that I will accomplish them	When facing difficult tasks, I am certain that I will achieve them
NGSE3: In general, I think that I can obtain outcomes that are important to me	In general, I think I can achieve things that are important to me
NGSE4: I believe I can succeed at most any endeavour to which I set my mind	I believe I can succeed at almost anything I set my mind to
TSY scale: 'Not true at all' and 'Really true'	Strongly disagree and Strongly agree
SPCCS introductory text: Below are 12 situations in which you might need to communicate. People's abilities to communicate effectively vary a lot and sometimes the same person is more competent to communicate in one situation than in another. Please indicate how competent you believe you are to communicate in each of the situations described below. Indicate in the space provided at the right of the item your estimate of your competence. Presume 0= completely incompetent and 100= completely competent.	Below are 12 situations in which you might need to communicate. Some people find it easier or harder to communicate in one situation over another. Please indicate how good you think you are at communicating in each of the situations described below by giving yourself a score out of 100 (where 0 = really bad and 100 = really good).

Analysis

Secondary analysis will be intention-to-treat (ITT), in which we test the hypothesis that being assigned a place on the programme has an effect on each secondary outcome.

Analysis of the effect on **English attainment** will be carried out using an OLS regression,

[12] See the following for more details: https://www.aqa.org.uk/subjects/english/gcse/english-language-8700/non-exam-assessment-administration2. [Last accessed: 25 October 2018].

$$Y_i = \beta_0 + \beta_1 \cdot T_i + X_i \cdot \Gamma + \epsilon_i$$

where:

- Y_i is the outcome for English GCSE performance, measured by the mean average point score[13] for English Language and Literature;
- T_i is a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i is a vector of individual-level stratification variables (School ID and EVERFSM_6_P) and the baseline attainment measured through KS2 Reading mark (KS2_READMRK); and
- ϵ_i is the individual-level error term. Standard errors will be corrected for heteroskedasticity.

Analysis of the effect on **Maths attainment** will be carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 \cdot T_i + X_i \cdot \Gamma + \epsilon_i$$

where:

- Y_i is the outcome for Maths GCSE performance, measured by point score;
- T_i is a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i is a vector of individual-level stratification variables (School ID and EVERFSM_6_P) and the baseline attainment measured through KS2 Maths mark (KS2_MATTOTMRK); and
- ϵ_i is the individual-level error term. Standard errors will be corrected for heteroskedasticity.

Analysis of the effect on **Self-efficacy** will be carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 \cdot T_i + X_i \cdot \Gamma + \epsilon_i$$

where:

- Y_i is the outcome for Self-efficacy, measured by the mean average score for the individual on the NGSE;
- T_i is a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i is a vector of individual-level stratification variables (School ID and EVERFSM_6_P) and the baseline attainment measured through separate KS2 Reading and Maths marks (KS2_READMRK and KS2_MATTOTMRK respectively); and
- ϵ_i is the individual-level error term. Standard errors will be corrected for heteroskedasticity.

[13] Whilst Uniform Mark Scale (UMS) scores provide a more granular measure, they are no longer available from the NPD, so point scores are used here.

(Academic attainment is known to be correlated with self-efficacy[14], so is included here as a covariate to increase the precision of our estimate).

Analysis of the effect on **Teamwork** will be carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 \cdot T_i + X_i \cdot \Gamma + \epsilon_i$$

where:

- Y_i is the outcome for Teamwork, measured by the mean average score for the individual on the TSY;
- T_i is a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i is a vector of individual-level stratification variables (School ID and EVERFSM_6_P); and
- ϵ_i is the individual-level error term. Standard errors will be corrected for heteroskedasticity.

Analysis of the effect on **Social confidence** will be carried out using an OLS regression,

$$Y_i = \beta_0 + \beta_1 \cdot T_i + X_i \cdot \Gamma + \epsilon_i$$

where:

- Y_i is the outcome for Social confidence, measured by the mean average score for the individual on the SPCCS[15];
- T_i is a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i is a vector of individual-level stratification variables (School ID and EVERFSM_6_P); and
- ϵ_i is the individual-level error term. Standard errors will be corrected for heteroskedasticity.

The evaluation protocol states that we will also run an ordered logit as a robustness check on the estimates for self-efficacy and teamwork (which are ordinal scales). However, we no longer believe that this is a sensible approach as we are using mean average scores from these scales to calculate the outcome, which effectively makes the outcome variable continuous. An ordered logit robustness check is therefore not included here.

Baseline academic attainment has not been used as a control variable in the analysis of teamwork or social confidence as there is no strong evidence, or theoretical argument, to suggest that these outcomes are correlated with attainment.

[14] Gutman, L. M., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. Education Endowment Foundation, London, 11.

[15] As specified above in 'Outcomes', effects will also be calculated for the sub-scores from the SPCCS, which relate to the different forms of communication assessed by the scale: 'Public', 'Meeting', 'Group', 'Dyad', 'Stranger', 'Acquaintance', and 'Friend'.

Subgroup analyses

We will conduct analysis on the primary outcome for the subgroup of pupils who are registered for free school meals in the NPD (using the EVERFSM_6_P variable), using the same model as our primary analysis, with the addition of an interaction between treatment assignment and FSM status, to assess whether there is a significant difference in the treatment effect between FSM students and others. The model we will use for this analysis is as follows:

$$Y_i = \beta_0 + \beta_1 \cdot T_i + X_i \cdot \alpha + Z_i \cdot \gamma + \delta \cdot T_i \cdot Z_i + \epsilon_i$$

where:

- Y_i is the outcome for overall GCSE performance, measured by capped average point scores across eight best GCSEs (Attainment 8);
- T_i is a binary indicator for the treatment assignment (1 if the pupil is assigned to treatment and 0 if not);
- X_i is a vector of the other individual-level stratification variable (school) and the baseline attainment measured through separate KS2 Reading (KS2_READMRK) and Maths (KS2_MATTOTMRK) marks;
- Z_i is a binary variable indicating the FSM status (EVERFSM_6_P) of individual *i*; and
- ϵ_i is the individual-level error term. Standard errors will be corrected for heteroskedasticity.

In this model, β_1 denotes the effect of the invention for participants who have not been eligible for FSM in the past 6 years (Z=0) and $\beta_1 + \delta$ denotes the effect of the intervention on those who have been eligible for FSM in past 6 years (Z=1). The difference in intervention effects between the two groups is therefore denoted by δ.

We will also estimate the treatment effect for the subsample of participants who have been eligible for FSM in the past 6 years (EVERFSM_6_P = 1) and compare this to the estimated treatment effect for those not eligible for FSM (EVERFSM_6_P = 0). This will be operationalised using the regression model in the primary analysis for each of these two groups.

Longitudinal analyses

As additional pieces of analysis, to be reported in Spring 2022 and Autumn 2023, we plan to investigate the impact of the intervention on a small number of longitudinal outcome measures. Whether we investigate the longitudinal outcome measures, will depend on the results of the initial analysis. If there are threats to the validity that invalidate reasonable longitudinal analysis, the approach outlined below will not be implemented[16]. The outcomes that we plan to investigate are:

- Participation in higher education
- Participation in the National Citizen Service
- Post-16 educational attainment
- Employment status

[16] See EEF's Guidance on Conducting Longitudinal Analyses. https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Writing_a_Protocol_or_SAP/longitudinal_guidance.pdf

Participation in higher education will be captured by a binary indicator for an individual having commenced higher education at any point in the first two academic years following the completion of compulsory education or training . While considering two years means we will have to wait longer for the data to become available, this is worthwhile given that a non-trivial number of students will take a gap year between finishing KS5 and starting university. Participation data will be derived from the commencement of studies date in the HESA dataset. Given that the students were in year 11 during the 2018/19 school year we will observe whether the individual registered at a university in either the 2022/23 or 2023/24 academic years[17].

Participation in the National Citizen Service (NCS) will be captured by a binary indicator for whether the individual has participated in NCS between the culmination of the Community Apprentice programme and the most recent NCS programme. This data will be obtained directly from NCS.

Post-16 attainment will be measured by a binary indicator for whether or not the individual had achieved a level 3 qualification by age 19.[18] This variable will be derived from a combination of NPD and Individualised Learner Record (ILR) [19] data. The ILR dataset covers further education colleges while the NPD dataset covers the remainder of post-16 educational institutions.

Employment status will be measured by a binary indicator, taking the value of one if the individual is either employed in the tax year containing the academic year that is two years after the completion of compulsory education or training. or has enrolled in university over the same period.[20] Employment information will be obtained from HMRC P45 data, this will be derived from a continuous variable in the P45 data which states the number of days for which an individual was employed over the tax year. University registration will be considered in the 2022/23 and 2023/24 academic years, while employment will be considered using the tax year ending April 2024.

The base category will be those not in education or employment. In order to be classed as employed, the individual must have been working for at least 50% of the tax year. To check whether this assumption affects the estimated impact of the intervention, the cut-off will be varied from 40% to 60% in ten percentage point increments to form a sensitivity analysis.

The model used to perform the analysis would be as follows:

$$Y_i = \beta_0 + \beta_1 \cdot T_i + X_i \cdot \Gamma + \epsilon_i$$

[17] This definition may miss out those individuals who take a break between compulsory training and higher education that is greater than one year, and those who spend a longer period of time studying pre-university.

[18] In 2017, 60.6% of individuals aged 19 in the UK had achieved a level 3 qualification. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/709682/L23_attainment_2018_main_text.pdf

[19] ILR data contains information on work-based learning and further education in England. This data is collected by institutions that are in receipt of Education and Skills Funding Agency (EFSA) funding.

[20] The implicit assumption here is that those who enrol at a university will continue their studies for a meaningful length of time.

Where, Y_i is the outcome under consideration which would be one of: participation in higher education, participation in the National Citizen Service, post-16 attainment, or employment status;

T_i is the binary indicator for having received the intervention;

X_i is a vector of the other individual-level stratification variable (school) and baseline attainment measured through separate KS2 Reading (KS2_READMRK) and Maths (KS2_MATTOTMRK) marks; and

ϵ_i is the individual-level error term. Standard errors will be corrected for heteroskedasticity.

While we are reasonably confident of our ability to access the data capturing participation in higher education, participation in the National Citizen Service and post-16 attainment. We are less sure on the availability of employment status since this data would either need to come from the new LEO data source or from HMRC data. As such we will aim to deliver all elements of the analysis but are uncertain as to whether the analysis of employment status is deliverable within the confines of the project.

Imbalance at baseline

We will assess imbalance at baseline, and for the sub-sample of those analysed, by calculating the following values in each case and cross-tabulating by treatment arm:

- Count and % male participants
- Count and % female participants
- Mean and standard deviation of Key Stage 2 Maths mark
- Mean and standard deviation of Key Stage 2 Reading mark
- % Ever 6 FSM

The difference in mean KS2 scores between the treatment and control group will be expressed as Hedges' g effect sizes.

Missing data

First, we will report the number of complete observations (those without any data missing). This could identify the following types of missing data:

- Missing pre-treatment covariates
- Missing outcome data

Data can be missing completely at random (MCAR), missing at random (MAR), or missing not a random (MNAR). These data types and forms of missingness are summarised in Table 1 and addressed in more detail below. In line with EEF guidelines on resource allocation, any imputation will be restricted to the primary analysis and will only be carried out when more than 5% of the data is missing. Schultz and Grimes[21] suggest that, when less than 5% of data is missing, there is likely to be little bias introduced to estimated treatment effects, so we have adopted this threshold here.[22]

Type of missing variable	Type of missing data	Further analysis

[21] Schulz, K. F., & Grimes, D. A. (2002). Sample size slippages in randomised trials: exclusions and the lost and wayward. *The Lancet*, *359*(9308), 784.
[22] This is also in line with the convention provided in EEF's guidance on statistical analysis.

Pre-treatment covariates	MAR	Multiple imputation before analysis of treatment effects
Pre-treatment covariates	MNAR	Sensitivity analysis
Outcome data	NA	None. Missing outcome data will not be imputed and these observations will therefore be lost to analysis.

Missing pre-treatment covariates

All observations with missing pre-treatment covariates will be included in the analysis as long as the outcome measure and treatment assignment are not missing. We know that we have complete data for School ID and EVER6_FSM_P. It is likely, however, that some observations will be missing values for baseline attainment (KS2 Reading and KS2 Maths marks) due, for example, to pupils in the sample who did not sit the test(s). For any missing data of this type (where more than 5% of the data is missing), we will first try to establish which variables are predictive of the missing data. To do this, we will create a new variable that is a binary indicator of missingness and look for its predictors using a logistic regression model to establish correlations with other the variables in the dataset. Missing KS2 Reading data will be modelled as follows:

$$M_i \sim binomial(p_i);\ logit(p_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + {X_3}' \cdot \beta_3$$

where:

- M_i is the binary variable for missingness (equal to 1 if missing and 0 if not missing);
- p_i is the probability that a given observation is missing the KS2 Reading mark;
- x_1 is the EVER6_FSM_P variable;
- x_2 is the KS2 Maths mark; and
- ${X'}_3$ is a vector of binary variables for the School ID[23].

The same model will used to model missingness for the KS2 Maths score, substituting KS2 Maths for KS2 Reading in the specification above. Where both KS2 Maths and KS2 Reading scores are missing, the following model will be used,

$$M_i \sim binomial(p_i);\ logit(p_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

where:

- M_i is the binary variable for missingness (equal to 1 if missing and 0 if not missing);
- p_i is the probability that a given observation is missing both KS2 Reading and KS2 Maths scores;
- x_1 is the EVER6_FSM_P variable; and
- x_2 is the School ID.

If the coefficients in the regression are significant (i.e. the values are missing conditional upon other variables in the model) and missingness does not depend on unobserved covariates, imputation will provide an unbiased estimate of the true values. Multiple

[23] NB: As School ID is a categorical variable with 30 categories, an F-Test will be used to test for the overall significance of school ID.

imputation (MI) will be carried out using the Markov chain Monte Carlo (MCMC) method to predict the missing values prior to the analysis of treatment effects. We will then estimate the treatment effect using the imputed data and compare our result with the primary analysis (conducted on complete cases only).

If the point estimates of these two datasets are not similar (more than 0.05 standard deviations apart), it is likely that the data is MNAR and sensitivity analysis will be carried out. This will entail modelling missingness as above with all possible combinations of the variables available (EVER6_FSM_P, KS2 English and KS2 Maths scores, and School ID).

If, after modelling missingness, as described above, it is found that our covariates do not explain the missingness, this will imply that the data is either MCAR or MNAR. In this case, we will be conservative and assume that the data is MNAR and conduct sensitivity analysis. These sensitivity analyses will investigate the sensitivity of the point estimate of the treatment effect to changes in model specification (and hence sample definition), through the inclusion and exclusion of variables for which observations are missing, as well as using null imputation to provide a more intuitive analysis based on a full sample of data.

Missing outcome data

No intermediate outcomes have been identified as appropriate to use to impute primary outcome data; the secondary outcomes in this study (teamwork, communication confidence and self-efficacy) are not sufficiently predictive of the primary outcome. Observations with missing primary outcome data will therefore be dropped from the analysis and a complete case analysis will be run.

Compliance

We will undertake two types of exploratory analysis relating to compliance.

First, in the case of one-sided non-compliance (where some individuals assigned to treatment do not participate), we will use an instrumental variables approach to estimate the Complier Average Causal Effect (CACE). In the context of the trial, to be considered as minimally compliant with the treatment, a student must have attended at least the following combination of sessions:

1. 5 weekly sessions
2. 1 business mentoring session

This minimal dosage was estimated by the delivery organisation (Envision). It is important to note that we do not know the true minimal amount of dosage needed to generate a treatment effect, so the cut-off chosen for minimal compliance is the delivery team's best estimate. The instrumental variable that we will use is treatment assignment, which is assumed to influence whether you participate in the programme but not the outcome variable in its own right.

Two key assumptions need to hold for this approach:

1. Being assigned to the treatment increases participation in the treatment. In this instance, students may only participate in the programme if they are assigned to treatment. We believe we will have sufficient participation among treatment group students for this assumption to hold. There is no ability to participate in the programme outside of the assignment to the treatment group.

2. Random assignment does not in itself impact outcomes. We have no reason to believe that the offer of the programme would improve attainment on its own, but instead believe the impact on attainment is achieved through participation in the programme.

The CACE estimation will use a two-stage least squares (2SLS) approach:

$$T_i = \gamma_0 + \gamma_1 Z_i + X_i\delta + u_i (1)$$

$$Y_i = \beta_0 + \beta_1 \hat{T}_i + X_i\alpha + \epsilon_i (2)$$

where:

- Z_i is a binary indicator for the treatment assignment (1 if the student is assigned to treatment and 0 if the student is assigned to control);
- T_i is whether a student met the minimal compliance threshold;
- X_i is a vector of individual-level stratification variables (school and EVERFSM_6_P) and the baseline attainment measured through separate KS2 Reading (KS2_READMRK) and Maths (KS2_MATTOTMRK) marks;
- u_i is the error term in the first stage;
- ϵ_i is the error term in the second stage;
- $\hat{T}_i$ are the predicted levels of compliance with the programme from (1); and
- Y_i is the outcome for the overall GCSE performance, measured by capped average point scores across eight best GCSEs (Attainment 8).

Our second exploratory analysis will include a descriptive dosage analysis, in which we explore the relationship between dosage and our outcome variables. This analysis will not be causal, because dosage will be driven by unobserved factors such as motivation, so that although whether or not someone participates in the programme (at all) is influenced by the randomly assigned treatment variable, the decision within the treatment group to attend 4 classes rather than 6 is not random. The analysis will aim to provide useful evidence on how attendance relates to the programme's intended outcomes and may be useful for developing participant recruitment and retention goals.

To begin with, we will generate graphs plotting dosage against each of the outcome variables to explore the nature of the relationship. For example, we may find a linear relationship that shows a steady increase in attainment as dosage increases, or we may find a non-linear relationship that shows little impact on attainment at lower dosage levels and a marked increase at higher dosage levels.

If we observe a large increase in our outcome variable between dosage levels, we will test for whether that increase is significant using the Chow test. The Chow test entails splitting the data at the point where we observe the change and running separate regressions on each half of the split data. We will then compare the coefficients of the regressions to determine whether they are significantly different using an F-test. While these changes in the outcome variable cannot be fully attributed to the increased dosage, we will also explore how other observable characteristics change at various dosage levels. This would provide indicative evidence of areas of future research into the optimal dosage for the intervention.

If visual inspection of the graphed relationships between dosage and our outcome variables warrants it, we will attempt to model these relationships formally through the appropriate regression analyses.

Intra-cluster correlations (ICCs)

NA – as randomisation was done at the individual level.

Effect size calculation

Effect sizes will be expressed in terms of Hedges' g, using the following formula:

$$ES = \frac{M_1 - M_2}{SD^*_{pooled}}$$

Where,

- M_1 is the mean value of the outcome in the control group; and
- M_2 is the mean value of the outcome in the treatment group.

And,

$$SD^*_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

Where,

- n_1 is the number of observations analysed in the control group;
- n_1 is the number of observations analysed in the treatment group;
- SD_1 is the standard deviation of the outcome variable in the control group; and
- SD_2 is the standard deviation of the outcome variable in the treatment group.

The difference in means for each outcome variable will be recovered from the relevant regression equation above ($\hat{\beta}_1$).

Appendices

Appendix 1: Analysis code

In this appendix, we provide indicative analysis syntax to implement the models specified in the Statistical Analysis Plan using Stata 14. Eventual syntax may have small changes (e.g. variable name changes) that do not affect the syntax's implementation of the models specified above. Variables are as specified in the statistical analysis plan.

Primary outcome analysis

```
regress att8 assigned school_id everfsm_6_p readmark mathsmark, vce(robust)
```

is a linear regression model estimated on the full randomised sample data where `att8` is the Attainment 8 score (corresponding to Y in the regression equation), assigned is the treatment assignment, `school_id` is a school identifier, `ever6_fsm_p` is the FSM status of the individual, `readmark` is the KS2 English Reading mark and `mathsmark` is the KS2 Maths mark.

Secondary outcome analysis

The same syntax will be used for the secondary analysis as the primary analysis, replacing the variable names as appropriate.

Subgroup analysis

```
regress att8 assigned school_id readmark mathsmark everfsm_6_p assigned#
everfsm_6_p, vce(robust)
```

is a linear regression model estimated on the full randomised sample data where `att8` is the Attainment 8 score (corresponding to Y in the regression equation), assigned is the treatment assignment, `school_id` is a school identifier, `ever6_fsm_p` is the FSM status of the individual, `readmark` is the KS2 English Reading mark and `mathsmark` is the KS2 Maths mark.

CACE analysis

```
ivregress 2sls att8 comply school_id everfsm_6_p readmark mathsmark (comply
= assigned), vce(robust)
```

is an instrumental variables (two stage least squares) regression model estimated on the full randomised sample data where `att8` is the Attainment 8 score (corresponding to Y in the regression equation), `comply` is a binary indicator of compliance defined in the evaluation protocol, `assigned` is the treatment assignment, `school_id` is a school identifier, `ever6_fsm_p` is the FSM status of the individual, `readmark` is the KS2 Reading mark and `mathsmark` is the KS2 Maths mark.

Appendix 2: Secondary outcome survey

School Name:	Test Academy
Student Name:	Joe Bloggs
BIT ID:	XXYY

About this survey

This survey asks you to think about yourself and your abilities in a range of different ways. There are no right or wrong answers – just put what feels right for you. We won't share your individual scores with anyone.

Part 1

Read each statement below and circle a number on the scale. The scale goes from 1 to 5 (where 1 = 'Strongly disagree' and 5 = 'Strongly agree').

Survey Item	**Scale**				
	Strongly disagree	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
1. I will be able to achieve most of the goals that I have set for myself	1	2	3	4	5
2. When facing difficult tasks, I am certain that I will accomplish them	1	2	3	4	5
3. In general, I think that I can achieve things that are important to me	1	2	3	4	5
4. I believe I can succeed at most any endeavour to which I set my mind	1	2	3	4	5
5. I will be able to successfully overcome many challenges	1	2	3	4	5
6. I am confident that I can perform effectively on many tasks	1	2	3	4	5
7. Compared to other people, I can do most tasks very well	1	2	3	4	5

8. Even when things are tough, I can perform quite well	1	2	3	4	5

Part 2

Think about the times when you work in a team (either in or out of school). Read each statement below and circle a number on the scale. The scale goes from 1 to 5 (where 1 = 'Strongly disagree' and 5 = 'Strongly agree').

Survey Item	Scale				
	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1. I think that team-work is important	1	2	3	4	5
2. People who work in teams can learn more if they work by themselves	1	2	3	4	5
3. I feel confident in my ability to work in a team	1	2	3	4	5
4. I know how to give my team members feedback that will not hurt their feelings	1	2	3	4	5
5. I ask others for feedback	1	2	3	4	5
6. I make an effort to include other members of my group	1	2	3	4	5
7. I value the contributions of my team members	1	2	3	4	5
8. I treat my team members as equal members of the team	1	2	3	4	5
9. I am good at communicating with my team members	1	2	3	4	5
10. I feel confident in my ability to be a leader	1	2	3	4	5
11. I worked well with my business mentors	1	2	3	4	5

12. I worked well with my Envision coach	1	2	3	4	5

Part 3

Below are 12 situations in which you might need to communicate. Some people find it easier or harder to communicate in one situation over another. Please indicate how good you think you are at communicating in each of the situations described below by giving yourself a score out of 100 (where 0 = really bad and 100 = really good).

Survey Item	**My Score (out of 100)**
1. Present a talk to a group of strangers	
2. Talk with an acquaintance	
3. Talk in a large group of friends	
4. Talk in a small group of strangers	
5. Talk with a friend	
6. Talk in a large meeting of acquaintances	
7. Talk with a stranger	
8. Present a talk to a group of friends	
9. Talk in a small group of acquaintances	
10. Talk in a large meeting of stranger	
11. Talk in a small group of friends	
12. Present a talk to a group of acquaintances	

Part 4

We'd like to know a bit about the extra-curricular clubs and activities that you do.

Survey Item	Answer	
1. Do you take part in any extracurricular clubs or activities? (This could be something like Young Enterprise, Scouts, Guides, Duke of Edinburgh, a drama club or a sports club).	Yes	No
2. If you answered 'Yes' to question 1, please write down the name of the activities here:		

Thank you!